EL623592121US

Express Mail EL623592121US

Attorney Docket No. 9432-000154

METHOD AND APPARATUS FOR ADAPTIVELY BINARIZING COLOR DOCUMENT IMAGES

FIELD OF THE INVENTION

[0001] The present invention relates to methods and apparatus for binarizing images, and more particularly to methods and apparatus for binarizing color or gray scale images under complex backgrounds.

BACKGROUND OF THE INVENTION

[0002] Optical character recognition (OCR) of black-and-white images is known. However, the popularity of color documents has created a need for text recognition of gray level and/or color characters, often with a complex background. For example, text with background of this type may often be found in advertisements and magazine articles. Sometimes, text is encountered that is on a complex textured background, or the background gradually changes from one color to another. This type of background is difficult to handle with traditional global thresholding methods.

[0003] More particularly, global thresholding methods are utilized in at least one current optical character recognition (OCR) software package. The generation of a single global threshold for an entire image is fast and simple. However, a global threshold provides satisfactory results only when an image has a highly even background. Even with user intervention, OCR software with

global thresholding cannot handle images with uneven illumination or complicated backgrounds such as a textured background.

SUMMARY OF THE INVENTION

[0004] One configuration of the present invention therefore provides a method for binarizing an image having *N* columns and *M* rows of pixels and a first column forming a first edge of the image, a last column forming a second edge of the image opposite the first edge, a first row of the image forming a third edge of the image and a last row of the image forming a fourth edge of the image opposite the third edge. The method, which produces an array of binarized pixels, includes:

[0005] (a) initializing, for each column of the image, a first variable representing a local column low pixel value and a second variable representing a local column high pixel value, and, for each row of the image, a third variable representing a local row low pixel value and a fourth variable representing a local row high pixel value;

[0006] (b) iteratively repeating steps (c) through (f) for each column of the image, from the first column to the last column;

[0007] (c) iteratively repeating steps (d) through (f) for each row of the image, from the first row to the last row;

[0008] (d) determining a threshold value dependent upon the first variable and the second variable at the column of the location index, and upon

the third variable and the fourth variable at the row of the location index, the location index being dependent upon the iterated column and the iterated row;

[0009] (e) comparing a value representative of an image pixel at the location index with the determined threshold value, and

[0010] (f) setting a binarization pixel for the location index to either a first value or a second value, dependent upon results of the comparison, and adjusting values of either the first variable and the third variable, or the second variable and the fourth variable dependent upon the results of the comparison.

[0011] Another configuration of the present invention provides a computing apparatus for binarizing an image having N columns and M rows of pixels and a first column forming a first edge of the image, a last column forming a second edge of the image opposite the first edge, a first row of the image forming a third edge of the image and a last row of the image forming a fourth edge of the image. The computing apparatus includes a memory and a processor operatively coupled to the memory for reading and storing values therein, and the computing apparatus is configured to:

[0012] (a) initialize in the memory, for each column of the image, a first variable representing a local low first direction pixel value and a second variable representing a local high first direction pixel value, and, for each row of the image, a third variable representing a local low second direction pixel value and a fourth variable representing a local high second direction pixel value;

[0013] (b) iteratively repeat (c) through (f) for each column of the image, from the first column to the last column;

[0014] (c) iteratively repeat (d) through (f) for each row of the image, from the first row to the last row;

[0015] (d) determine a threshold value dependent upon the first variable and the second variable at the column of the location index, and upon the third variable and the fourth variable at the row of a location index, the location index being dependent upon the iterated column and the iterated row;

[0016] (e) compare a value representative of an image pixel at the location index with the determined threshold value, and

[0017] (f) store, in the memory, a binarization pixel for the location index to either a first value or a second value, dependent upon results of the comparison, and adjust stored values of either the first variable and the third variable, or the second variable and the fourth variable dependent upon the results of the comparison,

[0018] wherein the iterations (b) and (c) produce an array of binarization pixels stored in the memory.

[0019] Yet another configuration of the present invention provides a machine readable medium or media having recorded thereon instructions configured to instruct a computing apparatus having a memory and a processor operatively coupled to the memory for reading and storing values therein to:

[0020] (a) initialize in the memory, for each column of an image having N columns and M rows of pixels and a first column forming a first edge of the image, a last column forming a second edge of the image opposite the first edge, a first row of the image forming a third edge of the image and a last row of the

image forming a fourth edge of the image, a first variable representing a local low first direction pixel value and a second variable representing a local high first direction pixel value, and, for each row of the image, a third variable representing a local low second direction pixel value and a fourth variable representing a local high second direction pixel value;

[0021] (b) iteratively repeat (c) through (f) for each column of the image, from the first column to the last column;

[0022] (c) iteratively repeat (d) through (f) for each row of the image, from the first row to the last row;

[0023] (d) determine a threshold value dependent upon the first variable and the second variable at the column of the location index, and upon the third variable and the fourth variable at the row of a location index, the location index being dependent upon the iterated column and the iterated row;

[0024] (e) compare a value representative of an image pixel at the location index with the determined threshold value, and

[0025] (f) store, in the memory, a binarization pixel for the location index to either a first value or a second value, dependent upon results of the comparison, and adjust stored values of either the first variable and the third variable, or the second variable and the fourth variable dependent upon the results of the comparison,

[0026] wherein the iterations (b) and (c) produce an array of binarization pixels stored in the memory.

[0027] Further areas of applicability of the present invention will become apparent from the detailed description provided hereinafter. It should be understood that the detailed description and specific examples, while indicating the preferred embodiment of the invention, are intended for purposes of illustration only and are not intended to limit the scope of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] The present invention will become more fully understood from the detailed description and the accompanying drawings, wherein:

[0029] Figure 1 is a drawing of a flow chart illustrating one configuration of a method for binarizing a color image.

[0030] Figure 2 is a representation of the arrangement of pixels in an image, such as that used as input to the method represented in Figure 1.

[0031] Figure 3 is a simplified block diagram illustrating one configuration of a computing system suitable for performing the method illustrated in Figure 1.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0032] The following description of the preferred embodiment(s) is merely exemplary in nature and is in no way intended to limit the invention, its application, or uses.

[0033] Referring to Figure 1, one configuration 10 of the present invention embeds a self-learning process in the binarization of color or gray scale images under various complex backgrounds.

[0034] It has been found that, for many documents having complex backgrounds, the background of a document normally changes gradually as it is scanned, except for transitions from text to reversed text, and vice versa. If a pixel is in a darker area, the probability of the subsequent pixel being in a darker area is relatively higher as a result of correlation of color backgrounds in a surrounding area. Using this assumption, if a scan process is going through a low contrast area, the threshold for the next neighboring pixel is adjusted lower, as well.

[0035] Thus, in one configuration and referring to Figure 1, a document that has been scanned in color is converted to a gray scale image 12. For example, an RGB image (i.e., one in which each pixel is represented by an R (red) value, a G (green) value, and a B (blue) value) is converted to YIQ format. The YIQ_Y value representing luminance or gray scale value is used for binarization. (YIQ formats are known from the NTSC color television standard, in which "Y" is a perceived luminance signal, "I" is a color difference signal derived from R-Y, and "Q" is a color difference signal derived from B-Y, where "R" is a red signal and "B" is a blue signal. As used herein, the luminance signal or grayscale value is denoted YIQ_Y.)

[0036] In configurations utilizing grayscale rather than color images, no conversion 12 to YIQ is necessary, as the gray values of pixels are used directly.

[0037] For an image of N pixels in a first direction by M pixels in a second, perpendicular direction, memory locations for the following variables are assigned and initialized 14:

$$\begin{aligned} &X_{low}(i),\ i = 0, \ldots, N-1 \\ &X_{high}(i),\ i = 0, \ldots, N-1 \\ &Y_{low}(j),\ j = 0, \ldots, M-1 \\ &Y_{high}(j),\ j = 0, \ldots, M-1 \end{aligned} \tag{1}$$

where:

i is an index, ranging from 0 to N-1, of a column in the image;

j is an index, ranging from 0 to M-1, of a row in the image;

$X_{low}(i)$ is a local low column value;

$X_{high}(i)$ is a local high column value;

$Y_{low}(j)$ is a local low row value;

$Y_{high}(j)$ is a local high row value.

[0038] Figure 2 is an illustration showing an orientation of a rectangular image 100 showing the first column (column number 0) forming a first edge 102 of image 100 and the last column (column number N-1) forming a second edge 104 opposite first edge 102. Similarly, the first row (row number 0) forms a third edge 106 of image 100 and the last row (row number M-1) forms a fourth edge 108 opposite edge 106. This mapping is somewhat arbitrary, in that the rows and/or columns may be numbered in the reverse order, and the image may be rotated 90 degrees (i.e., the roles of the rows and columns may be interchanged) in either direction, as long as the resultant mapping is consistently used

throughout the method. However, for explanatory purposes, the mapping shown in Figure 2 will be assumed throughout.

[0039] In one configuration, initializing 14 the local variables is performed utilizing minimum and maximum values of luminosity *YIQ_Y* from the YIQ representation of the scanned image. Initialization 14 of the local variables is thus determined utilizing relationships written as:

$$\begin{aligned} X_{low}(i) &= YIQ\_Y_{min},\ i = 0, \ldots, N-1 \\ X_{high}(i) &= YIQ\_Y_{max},\ i = 0, \ldots, N-1 \\ Y_{low}(j) &= YIQ\_Y_{min},\ j = 0, \ldots, M-1 \\ Y_{high}(j) &= YIQ\_Y_{max},\ j = 0, \ldots, M-1 \end{aligned} \qquad (2)$$

where:

$$\begin{aligned} YIQ\_Y_{min} &= \mathrm{minimum}\{YIQ\_Y(i,j)\},\quad i = 0, \ldots, N-1,\quad j = 0, \ldots, M-1 \\ YIQ\_Y_{max} &= \mathrm{maximum}\{YIQ\_Y(i,j)\},\quad i = 0, \ldots, N-1,\quad j = 0, \ldots, M-1; \end{aligned} \qquad (3)$$

i.e., YIQ_Y_{min} is the minimum luminosity in the N by M image, YIQ_Y_{max} is a maximum luminosity in the N by M image, and $YIQ\_Y(i,j)$ is the intensity of a pixel of the image at an index *i* and an index *j*.

[0040] A set of nested loops is used to iterate over each pixel at a location index (*i,j*) in the YIQ-representation of the scanned image and to return 20 a binarized image when the iteration is complete. In the configuration represented in Figure 1, variables *i* and *j* are set 16 to zero, and a test 18 is performed to determine whether *i* has iterated over the entire width of the image. If it has, the iterations are complete, and a binarized image is returned 20. Otherwise, a test 22 is performed to determine whether *j* has iterated over the entire image height at the current index *i*. If it has, the *i* index is increased 24 and

another loop over *j* is performed, provided that *i* has not iterated 18 over the entire width of the image.

[0041] Otherwise, at the location (*i,j*), a determination 26 of a local threshold *T(i,j)* is made, utilizing a relationship written as:

$$T(i,j) = (X_{low}(i) + X_{high}(i) + Y_{low}(j) + Y_{high}(j))/4. \tag{4}$$

The Y-value *YIQ_Y(i,j)* at the corresponding location (*i,j*) is compared 28 to this local threshold. Thus, if:

$$YIQ\_Y(i,j) < T(i,j), \tag{5}$$

then 30:

$$\begin{aligned} &B(i,j) = 0 \\ &X_{low}(i) = (X_{low}(i) * w + YIQ\_Y(i,j))/(w+1) \\ &Y_{low}(j) = (Y_{low}(j) * w + YIQ\_Y(i,j))/(w+1) \end{aligned} \tag{6}$$

else 32:

$$\begin{aligned} &B(i,j) = 1 \\ &X_{high}(i) = (X_{high}(i) * w + YIQ\_Y(i,j))/(w+1) \\ &Y_{high}(j) = (Y_{high}(j) * w + YIQ\_Y(i,j))/(w+1) \end{aligned} \tag{7}$$

where:

* (asterisk) represents multiplication,

B(i,j) is the determined binarized image pixel at location index (*i,j*) that is stored in memory; and

w is a parameter.

[0042] In one configuration, a *B(i,j)* value of 0 is mapped to black and a value of 1 is mapped to white. However, in another configuration, a different, but consistent mapping is applied.

[0043] Threshold $T(i,j)$ adaptively changes as the image is scanned, as will be appreciated by observing that changes in either $X_{low}(i)$ and $Y_{low}(j)$, or in $X_{high}(i)$, and $Y_{high}(j)$ the occur, depending upon the consequences 30, 32 of each threshold comparison 28. Also, because of the updates made to $X_{low}(i)$, $Y_{low}(j)$, $X_{high}(i)$, and $Y_{high}(j)$ during binarization of the image, their values at any particular (i,j) pixel location do not necessarily represent actual minimum and maximum values of luminosity, either globally or locally.

[0044] Parameter w in one configuration is a user-adjustable parameter that may be thought of as defining a "localization region" for $X_{low}(i)$, $Y_{low}(j)$, $X_{high}(i)$, and $Y_{high}(j)$. However, parameter w is not required to adjustable in all configurations of the present invention. In one configuration, parameter w is made dependent upon image resolution. Those skilled in the art will recognize that the changes to $X_{low}(i)$, $Y_{low}(j)$, $X_{high}(i)$, and $Y_{high}(j)$ represent an operation utilizing a computational kernel. The kernel described by the equations above depends only on the current location index values of i and j, but in other configurations, other kernels are utilized that include dependencies on weighted values of X_{low}, Y_{low}, X_{high}, and Y_{high} at additional rows or columns, such as adjacent rows and columns.

[0045] The more pixels that are processed, the more reliable threshold $T(i,j)$ becomes for binarization. The reliability of values of $X_{low}(i)$, $Y_{low}(j)$, $X_{high}(i)$, and $Y_{high}(j)$ for determining each value of $T(i,j)$ also increase.

[0046] To further enhance performance in one configuration of the present invention, in one configuration, after the initialization 14 of the local

variable but prior to the looping iterations (e.g., between steps 14 and 16 in Figure 1), a pre-training process is applied to variables $X_{low}(i)$, and $X_{high}(i)$, and variables $Y_{low}(j)$, and $Y_{high}(j)$. The following pseudo-code describes four separate pre-training procedures, where **A_1**, **A_2**, **A_3**, and **A_4** are labels for each procedure:

A_1: *for i=N_1 to i=N_2*

for j=M_1 to j=M_2

if $YIQ\_Y(i,j) < (X_{low}(i) + X_{high}(i))/2$

then $X_{low}(i) = (X_{low}(i) * w + YIQ\_Y(i,j))/(w+1)$

else $X_{high}(i) = (X_{high}(i) * w + YIQ\_Y(i,j))/(w+1)$

A_2: *for i=N_2 down to i=N_1*

for j=M_2 down to j=M_1

if $YIQ\_Y(i,j) < (X_{low}(i) + X_{high}(i))/2$

then $X_{low}(i) = (X_{low}(i) * w + YIQ\_Y(i,j))/(w+1)$

else $X_{high}(i) = (X_{high}(i) * w + YIQ\_Y(i,j))/(w+1)$

A_3: *for i=N_1 to i=N_2*

for j=M_1 to j=M_2

if $YIQ\_Y(i,j) < (Y_{low}(i) + Y_{high}(i))/2$

then $Y_{low}(j) = (Y_{low}(j) * w + YIQ\_Y(i,j))/(w+1)$

else $Y_{high}(j) = (Y_{high}(j) * w + YIQ\_Y(i,j))/(w+1)$

A_4: *for i=N_2 down to i=N_1*

for j=M_2 down to j=M_1

if $YIQ\_Y(i,j) < (Y_{low}(i) + Y_{high}(i))/2$

$$then\ Y_{low}(j) = (Y_{low}(j) * w + YIQ\_Y(i,j))/(w+1)$$

$$else\ Y_{high}(j) = (Y_{high}(j) * w + YIQ\_Y(i,j))/(w+1)$$

[0047] In one configuration of the present invention, pre-training is performed by performing all four pre-training procedures **A_1**, **A_2**, **A_3**, and **A_4**. In other configurations, two pre-training procedures are performed, namely, one procedure selected from procedures **A_1** and **A_2**, and another procedure selected from **A_3**, and **A_4**. (For example, in one such configuration, pre-training procedures **A_1** and **A_3** are performed.) Such configurations may, but need not offer a user a choice of which of the four different combinations of pre-training procedures are performed. In yet another configuration, none of the pre-training procedures **A_1**, **A_2**, **A_3**, and **A_4** is performed.

[0048] *M_1*, *N_1*, *M_2*, and *N_2* define the size of an area in which initial training is performed, and:

$$\begin{aligned} &0 \le M\_1 \le M\_2 \le (M-1) \text{ and} \\ &0 \le N\_1 \le N\_2 \le (N-1). \end{aligned} \tag{8}$$

(By convention, for loops in which bounds *M*_1 and *M*_2, or *N*_1 and *N*_2 are equal, the loop is executed once.)

[0049] In configurations using any of pre-training procedures **A_1**, **A_2**, **A_3**, and **A_4**, pre-training is performed over a rectangular subset of the image, which may be, but need not be, the entire image. When the subset is large, more training or learning for parameters $X_{low}(i)$, $Y_{low}(j)$, $X_{high}(i)$, and $Y_{high}(j)$ is achieved. Values of *M_1*, *N_1*, *M_2*, and *N_2* in one configuration of the present invention are selected in accordance with a desired computational speed, because larger pre-training areas require greater computational time.

[0050] In one configuration of the present invention and referring to Figure 3, a computing apparatus 200 suitable for performing the methods disclosed herein is provided. Computing apparatus 200 includes a processor and a memory operatively coupled to the processor. Neither the memory nor the processor is shown, in Figure 3 but both are well known to those skilled in the art, as are techniques for operatively coupling the processor to the memory. The processor operates on images and variables (or arrays of variables) in the memory, and is able to store variables (or arrays of variables) in the memory or read variables (or arrays of variables) from the memory. Computing apparatus 200 also has a device configured to read instructions from an external machine-readable medium or media 206 and a scanner 204 for scanning images. In one configuration, instructions configured to instruct computing apparatus 200 to performing one or more configurations of the methods disclosed herein are recorded on medium or media 206.

[0051] Unlike methods having a predetermined threshold, configurations of the present invention utilize self-learning as the background of the image changes. Within the self-learning process, existing knowledge is accumulated and used iteratively. A threshold adjusts itself in one configuration as the process proceeds through rows and columns of a pixelized image. Therefore, configurations of the present invention work well even with uneven or textured backgrounds. In one configuration, the process trains itself, utilizing pixels of the image that have already been traversed. Resulting binarized

images are particularly suitable for optical character recognition (OCR) purposes, and are processed using OCR at least one configuration of the present invention.

[0052] In yet another configuration of the present invention, binarization is performed in "real time," i.e., during scanning of an image. This configuration is similar to the configuration shown in Figure 1 and described above, except that rather than initializing $X_{low}(i)$, $Y_{low}(j)$, $X_{high}(i)$, and $Y_{high}(j)$ as in equations 2 and 3 above, $X_{low}(i)$ and $Y_{low}(j)$ are initialized to the minimum possible pixel luminosity value and $X_{high}(i)$, and $Y_{high}(j)$ are initialized to the maximum possible pixel luminosity value. (For example, one configuration in which all luminosity values within an 8-bit integer value range are possible has a minimum possible luminosity value of 0 and a maximum possible luminosity value of 255.) In addition, the loop over variable *i* beginning at step 18 of Figure 1 is be performed as each scan line of the image is acquired. No pre-training is performed, however, because the image is not available for pre-training until binarization has already occurred.

[0053] In the configurations of the present invention described above, the luminance or gray value of each image pixel is utilized for binarization. However, it is possible to consistently substitute another value (for example, an R value from an RGB representation of a pixel, or a Q value from the YIQ representation of the pixel) for the luminance or gray value in configurations tailored for special purposes.

[0054] The description of the invention is merely exemplary in nature and, thus, variations that do not depart from the gist of the invention are intended

to be within the scope of the invention. Such variations are not to be regarded as a departure from the spirit and scope of the invention.